

# Integration of R to VTK, Adding Statistical Computation to a Visualization Toolkit

Jeff Baumes<sup>1</sup>, Andinet Enquobahrie<sup>1</sup>, Thomas O'Connell<sup>2</sup>, Tom Otahal<sup>3</sup>, Philippe Pébay<sup>4</sup>, Wesley Turner<sup>1\*</sup>, Michelle Williams<sup>5</sup>

1. Kitware Inc., Clifton Park, NY, U.S.A.

2. The Hamner Institutes for Health Sciences, 6 Davis Dr. P.O. Box 12137, Research Triangle Park, NC 12137, U.S.A.

3. Sandia National Laboratories, Albuquerque NM, U.S.A.

4. Sandia National Laboratories, MS 9159, PO Box 969, Livermore CA 94551, CA, U.S.A.

5. University of Washington, 1959 NE Pacific St., Seattle WA 98195 U.S.A.

\*Contact author: wes.turner@kitware.com

**Keywords:** R, VTK, Information Visualization, Informatics

Conveying the sense of complex data to the human mind requires sophisticated visualization methods. The Titan [4] informatics toolkit, a Sandia funded collaboration between Sandia National Laboratory and Kitware, represents an effort to add graphical, tabular, and geospatial visualization algorithms to the Visualization Toolkit (VTK) [3]. VTK is an open-source, freely available software system for 3D computer graphics, image processing and visualization. The Infovis additions to VTK expand the the toolkit to include visualization of spatially ambiguous entities. However, simply displaying relationships among entities is not sufficient. Statistical analysis such as that provided by R is a powerful tool for suppressing noise in the data and enhancing real relationships. This abstract describes the addition of an R interface to the VTK toolkit and introduces the use of the R engine in several VTK Infovis application areas.

The interface to R primarily consists of three VTK classes: `vtkRAdapter`, `vtkRInterface`, and `vtkRCalculatorFilter` that provide a separation between the VTK and R code. `vtkRAdapter` handles the conversions between R and VTK data structures allowing VTK tables and arrays to be converted to and from R SEXP data structures. `vtkRInterface` launches an instance of the R interpreter and manages access. `vtkRCalculatorFilter` uses instances of the `vtkRAdapter` and `vtkRInterface` classes to specify the R command to be executed. It also provides the programmatic interface to specify VTK data structures as inputs to the R analysis and R variables as outputs from the analysis.

We are currently working on two applications that leverage this interface. The first application is statistical hypothesis testing. More specifically, an option to perform statistical tests has been added to the descriptive [1] and contingency [2] VTK statistics engines: respectively, the Jarque-Bera normality and  $\chi^2$  independence tests. Although the test statistics themselves can and are directly computed in VTK, the calculation of the corresponding p-values requires that one-tailed probability values of the  $\chi^2$  distribution be available. We integrated this feature with the `EvalScript()` method of the VTK/R interface. Examples that demonstrate this functionality are available in the toolkit. A second application is an 'omics analysis viewer targeting the exploitation of genomic, metabolomic and proteomic data in the discovery and analysis of disease processes. Investigators at the University of North Carolina and the University of Washington, are collaborating with Kitware on this application to analyze genes, metabolites and proteins from biological assays; discover the statistically important components; and explore biological pathways modeling the disease process. This application links the visualization capabilities of VTK with R; allowing for statistical analysis including PCA, K-Means clustering, Student T tests, correlations, and pathway analysis. The combination of R with VTK allows for rapid prototyping of algorithms and for intuitive visualizations of the results.

## References

- [1] P. Pébay and D. Thompson, *Scalable descriptive and correlative statistics with Titan*, Sandia Report SAND2008-8260, Sandia National Laboratories, December 2008.
- [2] ———, *Parallel contingency statistics with Titan*, Sandia Report SAND2009-6006, Sandia National Laboratories, September 2009.
- [3] Bill Lorensen Will Schroeder, Ken Martin, *The visualization toolkit, an object oriented approach to 3d graphics*, third ed., Kitware, Inc., 2004.
- [4] Brian Wylie and Jeffrey Baumes, *A unified toolkit for information and scientific visualization*, Visualization and Data Analysis 2009 (Katy Borner and Jinah Park, eds.), SPIE, 2009, p. 72430H.